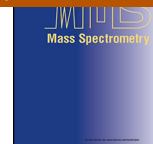




## International Journal of Mass Spectrometry

journal homepage: [www.elsevier.com/locate/ijms](http://www.elsevier.com/locate/ijms)

## A tale of a tail: Structural insights into the conformational properties of the polyglutamine protein ataxin-3

Charlotte A. Scarff<sup>a</sup>, Alessandro Sicorello<sup>a</sup>, Ricardo J.L. Tomé<sup>a,b</sup>, Sandra Macedo-Ribeiro<sup>b</sup>, Alison E. Ashcroft<sup>a,\*</sup>, Sheena E. Radford<sup>a,\*\*</sup><sup>a</sup> Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK<sup>b</sup> IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

## ARTICLE INFO

## Article history:

Received 15 June 2012

Received in revised form 22 August 2012

Accepted 23 August 2012

Available online 31 August 2012

## Keywords:

Ataxin-3

PolyQ

Amyloid

Ion mobility spectrometry

Electrospray ionisation-mass spectrometry

## ABSTRACT

Ataxin-3 is the protein responsible for the neurodegenerative polyglutamine disease Spinocerebellar ataxia type 3. Full structural characterisation of ataxin-3 is required to aid in understanding the mechanism of disease. Despite extensive study, little is known about the conformational properties of the full-length protein, in either its non-expanded healthy or expanded pathogenic forms, particularly since its polyglutamine-containing region has denied structural elucidation. In this work, travelling-wave ion mobility spectrometry–mass spectrometry and limited proteolysis have been used to compare the conformational properties of full-length non-expanded ataxin-3 (14Q) and its isolated N-terminal Josephin domain (JD). Limited proteolysis experiments have confirmed that the JD is stable, being extremely resistant to trypsin digestion, with the exception of the  $\alpha 2/\alpha 3$  hairpin which is flexible and exposed to protease cleavage in solution. The C-terminal region of ataxin-3 which contains the glutamine-rich sequences is largely unstructured, showing little resistance to limited proteolysis. Using ion mobility spectrometry–mass spectrometry we show that ataxin-3 (14Q) adopts a wide range of conformational states *in vitro* conferred by the flexibility of its C-terminal tail and the  $\alpha 2/\alpha 3$  hairpin of the N-terminal JD. This study highlights how the power of MS-based approaches to protein structural characterisation can be particularly useful when the target protein is aggregation-prone and has intrinsically unordered regions.

© 2012 Elsevier B.V. Open access under [CC BY](http://creativecommons.org/licenses/by/3.0/) license.

## 1. Introduction

## 1.1. Mass spectrometry as a structural biology tool

Over the last 20 years, mass spectrometry (MS) analysis of biological molecules has become an important component of the structural biology toolbox. With the advent of electrospray ionisation (ESI), the ability to ionise and analyse large biological molecules was soon realised and the use of MS in biological applications has become a powerful method of protein structure elucidation, especially for large and complex samples [1,2]. MS analyses benefit from being extremely rapid and having high sensitivity, requiring only picomolar amounts of sample.

Non-covalent MS analyses, where experiments are conducted under carefully controlled solvent and instrumental conditions, can

be used to yield information about a protein or protein complex that reflects its behaviour in solution. Consequently, non-covalent MS has been used to mass measure protein complexes  $\leq 10$  MDa [3–5] and to deduce their stoichiometry, topological arrangements and dynamics [6–10]. When ESI-MS analysis is coupled with ion mobility spectrometry (IMS) information on analyte shape, including the separation and size determination of co-populated protein conformers, as well as insights into the quaternary structure of macromolecular assemblies, can be gained simultaneously with mass measurements [11–18]. ESI-IMS-MS is particularly useful for the study of heterogeneous samples as it can separate and subsequently identify individual components within a mixture in a single experiment. It can also be used to track biomolecular reactions, yielding information about kinetic intermediates even when these are transient and/or lowly populated [17,19]. Such analyses are of prime importance when NMR spectroscopy or X-ray analysis of the biomolecule, or biomolecular complex, in question is unfeasible or extremely difficult.

Here, ESI-IMS-MS in conjunction with limited proteolysis studies have been used to provide insights into the structure and dynamics of ataxin-3, a clinically important polyglutamine-containing protein.

\* Corresponding author. Tel.: +44 113 343 7273; fax: +44 113 343 7273.

\*\* Corresponding author. Tel.: +44 113 343 3170.

E-mail addresses: [a.e.ashcroft@leeds.ac.uk](mailto:a.e.ashcroft@leeds.ac.uk) (A.E. Ashcroft), [s.e.radford@leeds.ac.uk](mailto:s.e.radford@leeds.ac.uk) (S.E. Radford).

## 1.2. Ataxin-3 and polyglutamine disease

Ataxin-3 is the protein associated with the neurodegenerative polyglutamine (polyQ) disease spinocerebellar ataxia type 3 or Machado–Joseph disease [20–23]. PolyQ diseases are a group of inherited neurodegenerative disorders, characterised by the accumulation of polyQ-containing neuronal aggregates in particular regions of the brain, whose causative agents are proteins with expanded polyQ stretches [24,25]. These polyQ-expanded proteins are a result of unstable expansion of CAG trinucleotide repeats in their respective gene products, with expansion beyond a particular length, usually >50 contiguous Q residues, resulting in disease [22,26].

An inverse correlation between the number of CAG repeats and the age of onset of disease has been observed for all the polyQ disorders and so polyQ aggregation and polyQ-directed protein misfolding are thought to play key roles in pathogenesis [22,27]. Despite these uniting features of polyQ disease, polyQ proteins show no sequence homology, except in their polyQ tracts, exhibit different functions, result in different pathognomonic symptoms and neurodegenerative profiles, and affect different regions of the brain [22,25,27].

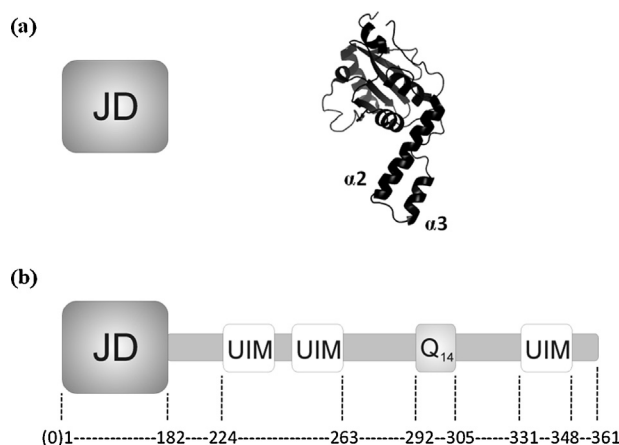
A detailed understanding of the mechanisms by which expression of polyQ-expanded proteins leads to pathology is required in order to generate therapeutic approaches against these disorders [25]. Characterisation of polyQ-containing protein structure, misfolding and aggregation states is an essential component of such studies. PolyQ-containing proteins, however, are particularly challenging to study as they are often structurally dynamic, contain intrinsically disordered regions and are aggregation-prone [22,28]. IMS–MS is therefore an ideal tool for the study of such systems yet, to date, has not been applied rigorously to the structural characterisation of this important set of disease-causing proteins.

## 1.3. Ataxin-3 structure

Ataxin-3 is a 42–52 kDa protein that contains an N-terminal Josephin domain (JD; residues 1–182) followed by two ubiquitin interacting motifs (UIMs), a polyQ stretch of variable length, and a variable C-terminal region [29]. The polyQ stretch consists of 12–44 glutamine residues in healthy individuals and results in disease when expanded beyond approximately 54 glutamine residues [30]. In one of the known isoforms of this protein, the predominant isoform in the brain, the C-terminal region includes a third UIM following the polyQ stretch (Fig. 1) [31]. Although the tertiary structure of the JD has been determined by use of solution NMR spectroscopy [29,32,33], little is known about the conformational properties of the full-length ataxin-3 protein, with its C-terminal region herein termed the “tail” [26].

The NMR structure of the JD revealed that this region consists of two sub-domains: a relatively rigid globular catalytic domain and a flexible  $\alpha 2/\alpha 3$  helical hairpin, the latter which exhibits fast local motions and behaves much as an exposed “waving hand” [29,32–34]. In contrast, the tertiary structure of full-length ataxin-3 has so far evaded detailed structural characterisation at the atomic level. Previous NMR and CD studies on ataxin-3 have suggested that the C-terminal region of ataxin-3 has no defined tertiary structure, yet has multiple secondary structural elements [32]. In addition, NMR data have indicated that the overall fold of the JD is not perturbed by the presence of the C-terminal region of ataxin-3 and that there are no significant interactions between the two regions [32].

To date, there has been a paucity of non-covalent MS experiments on polyQ-containing proteins, including ataxin-3. Indeed, a single ESI–MS study comparing the conformational properties in the presence of organic solvents or heat of the isolated JD with those of a construct of ataxin-3 containing the JD and the adjacent flexible



**Fig. 1.** Schematic diagrams of (a) the Josephin domain (JD), and (b) ataxin-3 (14Q). The cartoons illustrate the relative positions of the JD, the ubiquitin interacting motifs (UIMs) and the polyQ stretch (Q<sub>14</sub>). The NMR structure for the isolated JD (PDB 1YZB [29]) with the flexible  $\alpha 2/\alpha 3$  hairpin is highlighted as a ribbon diagram in (a).

region (183–291) prior to the polyQ stretch has been reported [35], with the authors concluding that the two constructs showed a similar stability to denaturation. Here, by the use of limited proteolysis and ESI–IMS–MS experiments we provide insights into the conformational properties of the non-expanded ataxin-3 (14Q), which contains a stretch of thirteen glutamine residues interrupted by a lysine residue after the first three glutamines, with a single glutamine some six residues later, and by convention is named 14Q, in comparison to the isolated JD. The results provide further evidence that the tail of non-expanded ataxin-3 does not contain defined tertiary structure nor does it interact with the JD. This work demonstrates the power of MS-based approaches to studying polyQ-containing proteins and proteins with intrinsically disordered regions.

## 2. Materials and methods

### 2.1. Protein expression and purification

The cDNAs for human ataxin-3 (14Q) (isoform 2) and the JD (ataxin-3 residues 1–182) were sub-cloned into pDEST17 plasmid vectors [36]. Upstream of the coding site, a sequence coding for an N-terminal hexahistidine tag and a linker region containing a cleavage site for the recombinant tobacco etch virus (rTEV) protease were incorporated. The proteins were expressed in the *E. coli* strain BL21(DE3)-pLysS and purified by nickel-affinity chromatography followed by histidine-tag cleavage with a hexahistidine-tagged rTEV protease [37]. Cleaved recombinant protein was separated from rTEV protease by nickel-affinity chromatography and ataxin-3 (14Q) and JD were then purified to homogeneity by gel-filtration chromatography using a Superdex S200 column. Each recombinant protein construct retains one additional glycine residue at the N-terminus due to the site of rTEV protease cleavage. Herein, this residue is thus numbered zero in the amino acid sequence to avoid confusion.

### 2.2. Sample preparation for MS

Protein samples were buffer exchanged into 10 mM ammonium acetate by use of ZEBRA desalting columns (Thermo Fisher Scientific, Loughborough, Leicestershire, UK) and prepared at a concentration of 10–20  $\mu$ M in 10 mM ammonium acetate for MS analysis. To perform limited proteolysis experiments, bovine trypsin (Sigma Aldrich, Gillingham, Dorset, UK) was added to each protein solution

in a 1:100 molar ratio and incubated at 37 °C for different time periods.

### 2.3. ESI-(IMS)–MS analysis

A Synapt HDMS quadrupole time-of-flight mass spectrometer (Micromass UK Ltd., Waters Corp., Manchester, UK) was used to analyse the JD, ataxin-3 (14Q) and limited proteolysis samples. Samples were introduced into the instrument by direct infusion nanoESI with in-house prepared gold-coated borosilicate glass capillaries. MS and IMS–MS spectra were recorded in nanoESI positive mode using the following instrument parameters: cone voltage 70 V; source temperature 60 °C; backing pressure 3.5 mbar; travelling wave height 9 V; travelling wave speed 350 m/s; IMS gas flow 20 mL/min. Data were processed by use of MassLynx v4.1 and Driftscope software supplied with the mass spectrometer. Estimated collision cross-sectional areas for different species were calculated by use of a non-covalent protein calibration of beta-lactoglobulin, avidin and concanavalin A, following the protocol described by Bush et al. [12]. Tryptic peptides resulting from limited proteolysis of the JD and ataxin-3 (14Q) were separated by IMS and further characterised by collision induced dissociation (CID) MS/MS sequencing when required.

### 2.4. Modelling

Model structures were generated in PyMOL (DeLano Scientific, San Carlos, CA, USA) by use of PDB structure 1YZB [29] as a starting template. Theoretical collision cross-sectional areas for PDB 1YZB and model structures were calculated by use of the MOBCAL algorithm projection approximation [38] corrected for protein shape using the linear fit to the projected superposition approximation (PSA) [39]. This approximation has been shown to be a better estimate of cross-section than the projection approximation as it takes into account the convex-ness or concave-ness of a protein structure [39].

## 3. Results and discussion

### 3.1. Non-covalent ESI-IMS–MS of the JD and ataxin-3

The extent of protein ionisation during the ESI process is correlated directly with the surface-exposed area and mass of the protein [40]. The lowest charge states detected within the  $m/z$  spectrum for a protein are most reflective of the native structure of that protein, whilst partially folded protein conformations exhibit intermediate charge states and denatured or unfolded proteins carry the highest number of charges. In the case of the JD, the expected average charge states, based on mass [40], under native conditions and denaturing conditions are 9+ and 22+, respectively. For ataxin-3 (14Q) these are 14+ and 36+. The ESI-MS spectra obtained for the isolated JD (measured mass 21,092.8 Da; calculated mass 21,093.0 Da) and ataxin-3 (14Q) (measured mass 41,309.0 Da; calculated mass 41,307.2 Da) are shown in Fig. 2. One narrow charge state distribution, corresponding to monomeric 6+ to 10+ charge state ions, was populated predominantly in the spectrum of the JD whereas a much broader range of charge states was observed in the spectrum of ataxin-3 (14Q). Both spectra were acquired under identical experimental conditions and comparison of the two spectra thus indicates that the JD adopts mainly a single, compact conformation whereas full-length ataxin-3 (14Q) adopts multiple conformations, some more compact (11+ to 15+) and others more extended (18+ to 38+). The observation of such a wide array of charge states for a protein analysed under native-like solvent conditions, as observed for ataxin-3 (14Q), is unusual but is typical for intrinsically disordered and semi-disordered proteins [41,42]. A second, very minor, charge state distribution (10+ to 17+) was also observed in the JD spectrum indicating the low population of a partially unfolded state (Fig. 2a, inset). It is interesting to note that a population of ions corresponding to a dimeric species was also observed in the spectrum of ataxin-3 (14Q). These were either not observed, or were only very weakly populated, in the spectrum of the isolated JD.

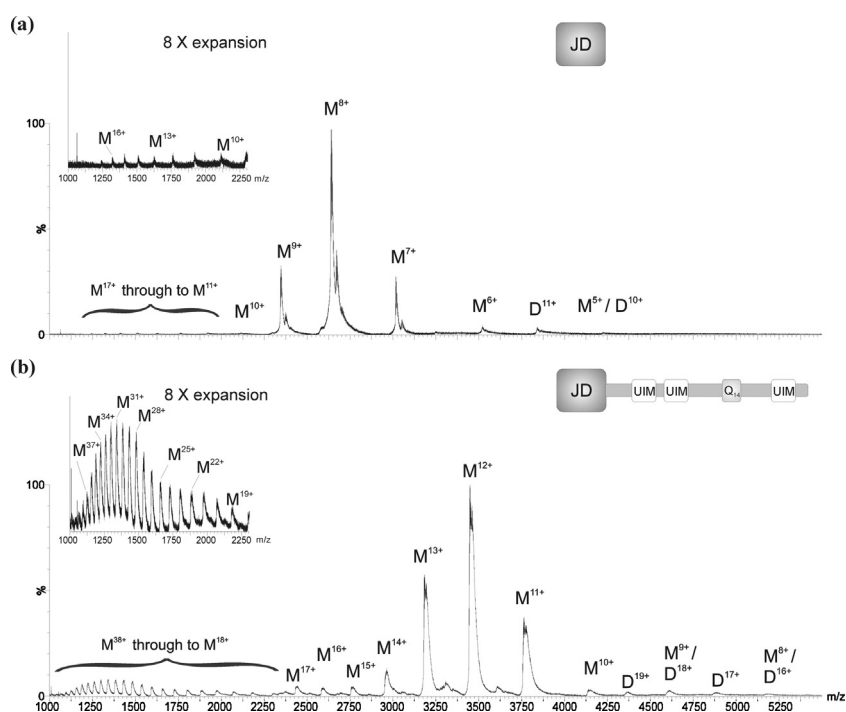
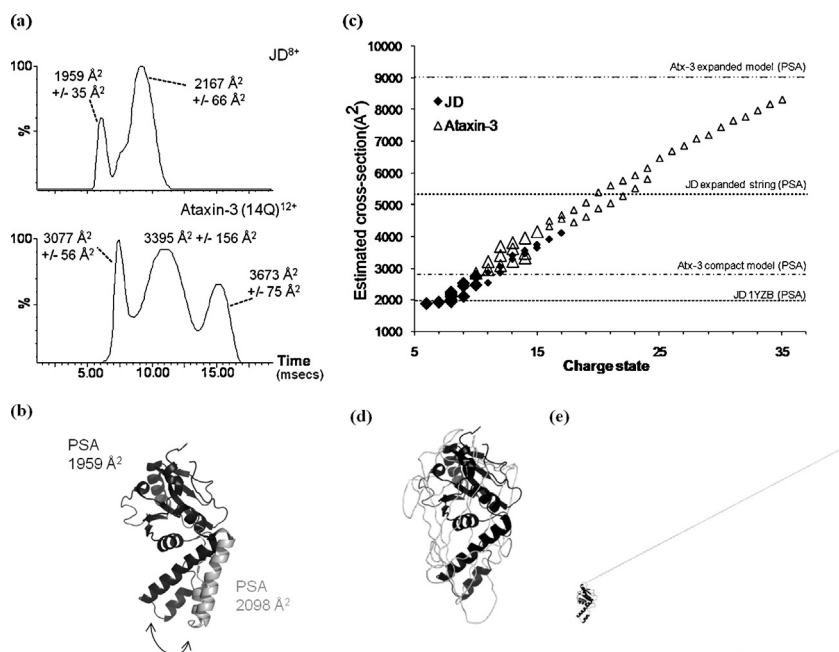


Fig. 2. ESI-MS spectra of (a) the isolated Josephin domain (JD) and (b) ataxin-3 (14Q). Insets show the expanded (8×) region  $m/z$  1000–2500. M, monomer; D, dimer.



**Fig. 3.** ESI-IMS-MS data for the JD and ataxin-3 (14Q). (a) Extracted arrival time distributions (ATDs) for the most abundant charge states observed within ESI-IMS-MS spectra for the JD (8+) (upper) and ataxin-3 (14Q) (12+) (lower). Estimated cross-sections for each of the main peaks observed in the ATDs, corresponding to maxima drift time values and their approximate dispersion values (based on peak width at half-height), are indicated (b). Schematic of PDB structure 1YJB (black) and modelled structure with  $\alpha 2/\alpha 3$  hairpin swung out (grey) overlaid; the calculated collision cross-sectional areas for these structures are 1959 and 2098 Å<sup>2</sup>, respectively. (c) ESI-IMS-MS estimated collision cross-sectional areas for all charge states observed for the isolated JD and ataxin-3 (14Q) (black diamond = JD, white triangle = ataxin-3 (14Q)). The size of symbols corresponds to the relative intensities of charge states observed within ESI-MS spectra. Theoretical calculations of cross-sectional areas from model structures are shown as horizontal dotted lines; PSA, projected superposition approximation [39]. Schematics illustrating (d) a compact model and (e) an expanded model of ataxin-3 (14Q).

ESI-IMS-MS experiments on ataxin-3 (14Q) support the notion that this protein adopts a range of compact and partially compact conformations that are co-populated in solution. Extracted arrival time distributions (ATDs) for the dominant charge states observed (in each case the maximum of the lowest charge state distribution which most reflects the native structure) for both the JD (8+) and ataxin-3 (14Q) (12+) exhibited multiple peaks over a range of drift times (Fig. 3a). Two main peaks (one with a shoulder) were observed in the extracted ATD for the dominant charge state of the JD with drift times of 6.16 and 9.23 ms (peak maxima values), from which collision cross-sectional areas of  $1959 \pm 35$  Å<sup>2</sup> and  $2167 \pm 66$  Å<sup>2</sup>, respectively, were estimated (Fig. 3a). These observations indicated the presence of at least two conformational families of the JD, differing in collision cross-sectional area by approximately 10%. The collision cross-sectional area calculated for the smaller of these two peaks is consistent with the theoretical value for the PDB structure 1YJB (1959 Å<sup>2</sup>). The collision cross-sectional area calculated for the second, more intense peak within the ATD is comparable with that which may be expected for the JD with its  $\alpha 2/\alpha 3$  hairpin swung out (2098 Å<sup>2</sup>), as modelled in Fig. 3b. The  $\alpha 2/\alpha 3$  hairpin of the JD domain is known to be solvent exposed and flexible in solution and to behave as an exposed “waving hand”, exhibiting fast local motions [29], and the ESI-IMS-MS results presented here are entirely consistent with this conclusion. In comparison, the broader array of conformations sampled by ataxin-3 (14Q) for the 12+ charge state ions differed in collision cross-sectional area by up to 20% ( $3077 \pm 56$  Å<sup>2</sup>,  $3395 \pm 156$  Å<sup>2</sup>, and  $3673 \pm 75$  Å<sup>2</sup>), Fig. 3a. The additional variation in conformational size exhibited by ataxin-3 (14Q) is likely to correspond to additional conformations adopted through movement and flexibility within the “tail”.

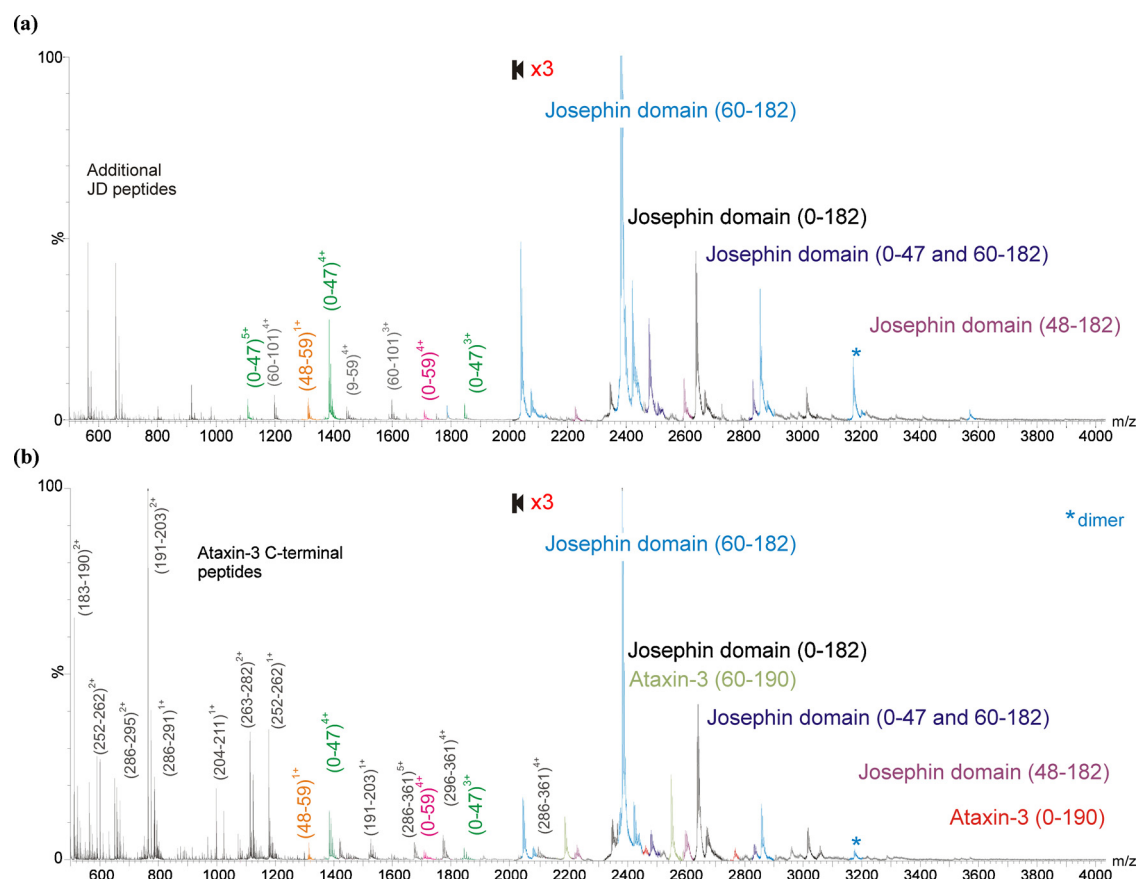
Ataxin-3 (14Q) species with estimated collision cross-sectional areas corresponding to a compact model where the protein exhibits a globular conformation (Fig. 3c and d), were observed for the

lowest charge states populated, which are most reflective of solution-phase structures. For higher charge states, representative of more expanded conformers, significantly larger collision cross-sectional areas were estimated from ion mobility measurements. Theoretical cross-sections for a fully extended model of the JD, with its amino acids extended in a linear string, and for ataxin-3 (14Q) with a folded JD but a fully extended C-terminal region (Fig. 3e), are shown in Fig. 3c to illustrate the relative range of cross-sections adopted.

The extent that the ESI-IMS-MS estimated collision cross-sectional areas for the partially folded conformers for the JD and ataxin-3 (14Q) mirror the conformational populations in solution is debatable. Gas-phase unfolding is driven by Coulombic repulsion forces. At low charge states, the effect of Coulomb repulsion between charges is negligible and attractive-folding forces, such as electrostatic interactions which are heightened in the gas phase, dominate [43,44]. Thus, the lowest charge states observed are most reflective of solution-phase structures. As the charge states increase, the effect of Coulomb repulsion becomes more dominant until at the highest charge states the increased Coulombic repulsion between charges forces the protein to adopt more open conformations [44]. Various intermediate conformations may form due to the interplay between attractive-folding and repulsive-Coulombic interactions [45]. For intrinsically disordered proteins, a scaling of overall size as a function of the net charge per residue has been observed [46].

Therefore, whilst some of the estimated collision cross-sectional areas are likely reflective of partially folded solution-phase structures (e.g., charge states 11+ to 15+ for the JD and charge states 16+ to 24+ for ataxin-3 (14Q), where the presence of multiple conformations indicates some degree of structure) the most expanded conformations may be a result of unfolding in the electrospray process due to Coulombic repulsion (charge states 16+ and 17+ for the JD and charge states > 25+ for ataxin-3 (14Q), where single protein





**Fig. 4.** ESI-MS spectra showing the major products of limited proteolysis for (a) the JD, and (b) ataxin-3 (14Q). In each case, spectra were acquired following the incubation of 10  $\mu$ M protein with 0.1  $\mu$ M trypsin in 10 mM ammonium acetate, pH 7.4, at 37  $^{\circ}$ C for 15 min. The protein sequences identified are indicated on the spectra.

conformations are observed exhibiting a uniform increase in cross-sectional area with increase in charge). It is clear, however, from the comparison between the spectra and collision cross-sectional area ESI-IMS-MS estimates for the JD and ataxin-3 (14Q) that ataxin-3 (14Q) adopts much more open conformations than the JD, strongly suggesting that the tail region of ataxin-3 is much less ordered than the JD as it is less restrained in the gas phase.

### 3.2. Limited proteolysis of the JD and ataxin-3

Both the JD and ataxin-3 (14Q) were subjected to limited proteolysis using the enzyme bovine trypsin to determine which regions of these proteins are accessible to this protease. Limited proteolysis is a powerful tool that can be used to gain additional structural information regarding a protein's solvent-exposed regions or its domain organisation. ESI-MS spectra obtained following the limited proteolysis of the JD and ataxin-3 (14Q) showed that many of the major products of limited proteolysis are the same for the corresponding regions of both species (Fig. 4). The major products of limited proteolysis observed for the JD are highlighted in Fig. 5a, and a comparison of the products obtained for the JD and the corresponding region of ataxin-3 (14Q) is presented in Table 1.

The key products observed following limited proteolysis of the JD resulted from cleavage of sites within the  $\alpha$ 2/ $\alpha$ 3 hairpin ( $\alpha$ -helices two and three cover residues 31–47 and 56–62, respectively) at amino acids Arg47 and Arg59. These two cleavage sites are known to be exposed in solution [29] and so proteolysis here was not unexpected. Nonetheless, the unique identification of these two residues as the most exposed sites to trypsin cleavage highlights the power of limited proteolysis to determine conformational properties of a target protein. The core of the JD (residues 60–182) was

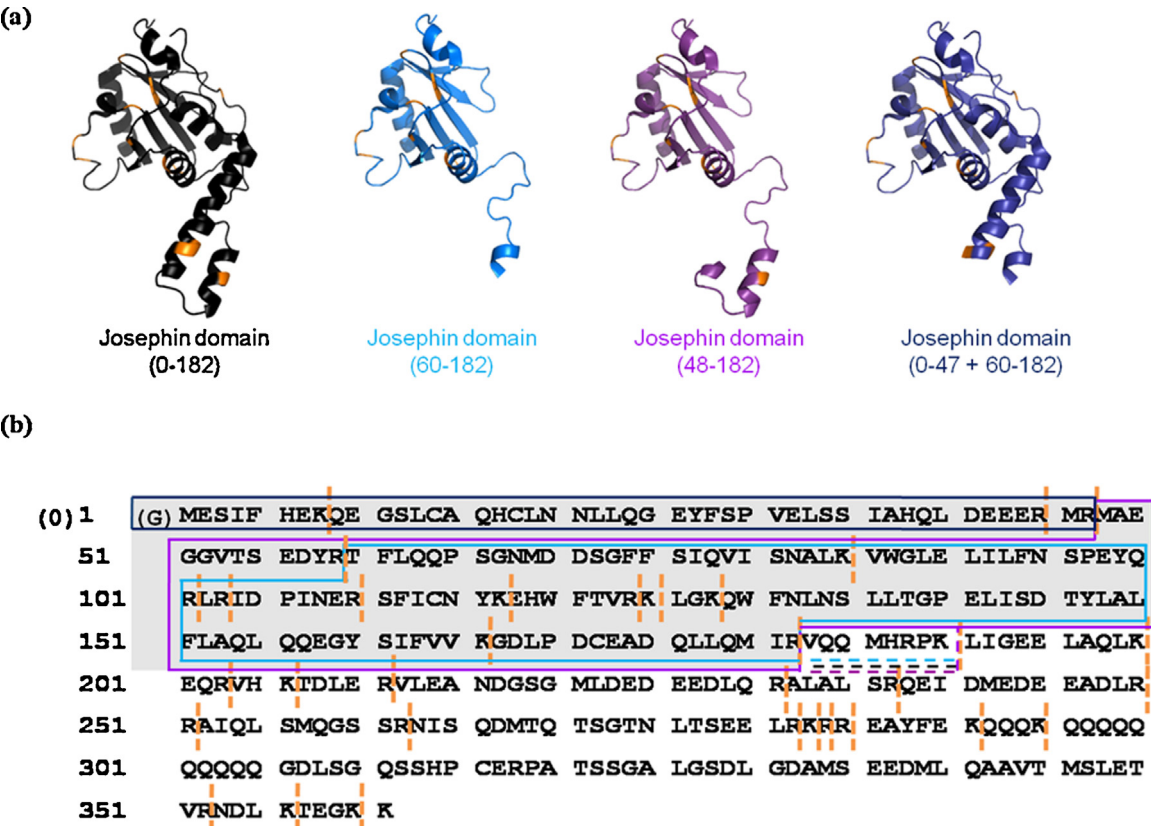
found to be highly resistant to proteolysis and was still present intact, even after 90 min of exposure to trypsin (data not shown). Interestingly, the experimental conditions were sufficiently gentle that a species corresponding to the 0–47 residue fragment coupled with the 60–182 residue fragment of the JD were observed in the spectra as a non-covalently bound complex; trypsin had cleaved the protein at residues 47 and 59 but the products of cleavage had not dissociated. The individual peptides produced by cleavage, 0–47, 0–59 and 48–59 were also observed (Fig. 4).

In limited proteolysis studies of ataxin-3 (14Q), the N-terminal region (residues 0–182) was found to undergo digestion at a similar rate to the isolated JD and produced the same products as seen for the JD alone (Figs. 4, 5 and Table 1). In addition, tryptic peptides consisting of residues 0–190, 48–190, and 60–190 were observed, resulting from cleavage at Lys190 (Tables 1, 2 and Fig. 5b).

**Table 1**

Major products resulting from the limited proteolysis of the Josephin domain (JD) and ataxin-3 (14Q). In each case, ESI-MS spectra were acquired following the incubation of 10  $\mu$ M protein with 0.1  $\mu$ M trypsin in 10 mM ammonium acetate, pH 7.4, at 37  $^{\circ}$ C for 15 min.

Major products of limited proteolysis (amino acid residues)	
Josephin domain	Ataxin-3 (14Q)
60–182	60–182
–	60–190
0–182	0–182
–	0–190
0–47 and 60–182	0–47 and 60–182
–	60–251
48–182	48–182
–	48–190



**Fig. 5.** Limited proteolysis of the isolated JD and ataxin-3 (14Q) following incubation of 10  $\mu$ M protein with 0.1  $\mu$ M trypsin in 10 mM ammonium acetate, pH 7.4, at 37 °C for 15 min. The major products of limited proteolysis of the JD domain (0–182, 60–182, 48–182, 0–47 + 60–182) are highlighted in (a) on JD structures in the form of ribbon diagrams, taken from the NMR structure for the isolated JD (PDB 1YZB [29]); these products resulted from cleavage at Arg47 and Arg59, both within the  $\alpha$ 2/ $\alpha$ 3 hairpin. The amino acid sequence for ataxin-3 (14Q) is shown in (b) with the major products of limited proteolysis (0–182, 60–182, 48–182, 0–47 + 60–182, 0–190 and 48–190). The JD domain (0–182) is highlighted in grey. The trypsin cleavage sites (orange bars) highlighted within the tail region (residues 183–361) all showed evidence of proteolysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

The observation of these products indicates that Arg182 (the final residue of the JD within the ataxin-3 (14Q) sequence) is buried, to some extent, within the tertiary structure of ataxin-3 (14Q), limiting access of trypsin to this cleavage site. This is consistent with work by Masino et al. on a similar construct, where equivalent cleavage products were observed for ataxin-3 (14Q) but the isolated

JD was not analysed for comparison in that study [26,32]. The generation of the same JD cleavage products in approximately the same intensities upon limited proteolysis of the JD and ataxin-3 (14Q) indicate that there are little or no differences in the accessibility of these cleavage sites in the absence or presence of the C-terminal “tail”. This suggests that the C-terminal region of ataxin-3 (14Q) does not interact with the JD in any significant manner, at least to the extent that the accessibility of Lys or Arg residues within the JD to trypsin cleavage is altered significantly.

Dimeric species of the product 60–182 were observed in the ESI mass spectra following limited proteolysis of both the JD and ataxin-3 (14Q), suggesting that the dimerisation domain of ataxin-3 (14Q) is contained within the core region of the JD. This is consistent with previous studies of ataxin-3 fibril formation, which indicated that protein aggregation initially involves self-association of the JD [36,47], and also with computer predictions which indicated the possibility of three main aggregation-prone regions of ataxin-3 residing within the Josephin domain: namely residues 73–96, 159–167 and 144–154 [35,48].

Peptides resulting from proteolysis at all possible cleavage sites within the ataxin-3 (14Q) C-terminal “tail” (residues 183–361) were identified within ESI-MS spectra after 15 min of trypsin exposure (Figs. 4, 5 and Table 2). The majority of cleavage sites within this region appear to be equally exposed, as peptides resulting from cleavage at these sites are observed instantaneously. This suggests strongly that the C-terminal region of ataxin-3 (14Q) possesses little, if any, persistent tertiary structure that protects its potential cleavage sites from trypsinolysis. Only two species corresponding to the presence of longer products within the C-terminal region,

**Table 2**  
Limited proteolysis cleavage products of the C-terminal region of ataxin-3 (14Q) digested with 0.01 (molar ratio concentration) bovine trypsin for 15 min at 37 °C.

Trypsin cleavage site (residue number)	Cleavage product observed	Mass-to-charge ratios of cleavage products observed
190	183–190	1023.6 (1+), 512.28 (2+)
200	191–200	557.3 (2+)
203	191–203	763.9 (2+)
206	191–206	946.0 (2+), 631.0 (3+)
211	204–211	997.5 (1+)
231	212–231	1118.0 (2+), 745.7 (3+)
237	232–237	630.4 (1+)
250	232–250	1102.5 (2+)
251	238–251	874.9 (2+)
262	252–262	1177.6 (1+), 589.3 (2+)
282	263–282	1113.0 (2+), 742.4 (3+)
283	263–283	1177.0 (2+)
284	263–284	837.0 (3+)
291	286–291	786.4 (1+)
295	286–295	649.8 (2+)
295	292–295	531.3 (1+)
352	353–361	1032.6 (1+)
356	357–361	562.3 (1+)
360	–	–

amino acid residues 286–361 and 296–361, were detected. These peptides incorporate the polyQ region in addition to the third UIM. However, there is no trypsin cleavage site between Lys296 and Arg353 within the C-terminal region so further proteolysis within this sequence is necessarily limited, and other peptides were also detected indicating that the trypsin cleavage sites at Arg352 and Lys356 in this region were indeed accessible for proteolysis. Unfortunately, therefore, it is not possible from these data to assess the compactness of the polyQ-containing region of ataxin-3 (14Q). Future work involving the study of ataxin-3 constructs with different proteases and/or polyQ stretches of different lengths will be needed to elucidate the conformational properties of this segment of the polypeptide chain, and its dependence on the polyQ length, in more detail.

#### 4. Conclusions

Here we have used ESI-IMS-MS and limited proteolysis experiments to gain new information regarding the conformational properties of a polyQ-containing protein, ataxin-3 (14Q). These experiments show, for the first time, that ataxin-3 (14Q) populates simultaneously a wide range of conformational states in solution, ranging from highly compact to widely extended conformers. Limited proteolysis experiments have indicated that the C-terminal “tail” region of ataxin-3 (14Q) has little resistance to trypsin digestion, consistent with this region being a dynamic, flexible entity able to adopt a wide repertoire of conformational states. However, the observation that the same tryptic peptides arise from the isolated JD and the corresponding JD within ataxin-3 (14Q) following limited trypsin proteolysis suggests that the C-terminal “tail” of ataxin-3 neither interacts with, nor shields, the JD from trypsinolysis. Furthermore, the dimerisation interface of ataxin-3 has been shown to be likely contained within the JD core region encompassing residues 60–182. Together, the results highlight the power of ESI-IMS-MS combined with limited proteolysis to examine the conformational properties of this fascinating class of proteins. Further work, focussing on ataxin-3 variants with glutamine repeats that extend to, and beyond, the pathogenic length, and the interaction of the protein with known modulators of ataxin-3 assembly, will cast new light into how and why these proteins assemble in a polyQ length-dependent manner. The ability of ESI-IMS-MS to analyse complex samples with a range of solution conditions, and to separate and structurally characterise lowly populated species within heterogeneous mixtures, will be essential to unravel the assembly mechanisms of these important protein systems.

#### Acknowledgements

CAS is funded by the Biotechnology and Biological Sciences Research Council (BBSRC; grant number BB/H024875/1). AS is a Wellcome Trust-funded postgraduate student (grant number 086866/Z/08/Z). The Synapt HDMS mass spectrometer was purchased with funds from the BBSRC's Research Equipment Initiative (BB/E012558/1). SMR's research is supported by Fundação para a Ciência e Tecnologia (FCT, Portugal) and FEDER (COMPETE/QREN) through grants PTDC/BIA-PRO/70455/2006 and PTDC/BIA-PRO/100059/2008. We are grateful to all members of the Ashcroft and Radford research groups for helpful discussions.

#### References

- [1] B.T. Ruotolo, C.V. Robinson, Aspects of native proteins are retained in vacuum, *Current Opinion in Chemical Biology* 10 (2006) 402–408.
- [2] C. Uetrecht, A.J. Heck, Modern biomolecular mass spectrometry and its role in studying virus structure, dynamics, and assembly, *Angewandte Chemie. International Ed. In English* 50 (2011) 8248–8262.
- [3] D.A. Shepherd, D. Velesler, J. Lichiere, A.E. Ashcroft, C. Cambillau, Unraveling lactococcal phage baseplate assembly by mass spectrometry, *Molecular and Cellular Proteomics* 10 (2011), M111.0009787.
- [4] C. Uetrecht, C. Verluise, N.R. Watts, P.T. Wingfield, A.C. Steven, A.J. Heck, Stability and shape of hepatitis B virus capsids in vacuo, *Angewandte Chemie. International Ed. In English* 47 (2008) 6247–6251.
- [5] M. Zhou, N. Morgner, N.P. Barrera, A. Politis, S.C. Isaacson, D. Matak-Vinkovic, T. Murata, R.A. Bernal, D. Stock, C.V. Robinson, Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding, *Science* 334 (2011) 380–385.
- [6] A.E. Ashcroft, Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes, *Natural Products Reports* 22 (2005) 452–464.
- [7] J.L. Benesch, B.T. Ruotolo, D.A. Simmons, C.V. Robinson, Protein complexes in the gas phase: technology for structural genomics and proteomics, *Chemical Reviews* 107 (2007) 3544–3567.
- [8] C. Bich, R. Zenobi, Mass spectrometry of large complexes, *Current Opinion in Chemical Biology* 19 (2009) 632–639.
- [9] A.J. Heck, Native mass spectrometry: a bridge between interactomics and structural biology, *Nature Methods* 11 (2008) 927–933.
- [10] J.A. Loo, Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrometry Reviews* 16 (1997) 1–23.
- [11] B. Bohrer, S.I. Merenbloom, S.L. Koeniger, A.E. Hilderbrand, D.E. Clemmer, Biomolecule analysis by ion mobility spectrometry, *Annual Review of Analytical Chemistry* 1 (2008) 293–327.
- [12] M.F. Bush, Z. Hall, K. Giles, J. Hoyes, C.V. Robinson, B.T. Ruotolo, Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology, *Analytical Chemistry* 82 (2010) 9557–9565.
- [13] A.C. Gill, K.R. Jennings, T. Wyttenbach, M.T. Bowers, Conformations of biopolymers in the gas phase: a new mass spectrometric method, *International Journal of Mass Spectrometry* 195/196 (2000) 685–697.
- [14] S.J. Hyung, C.V. Robinson, B.T. Ruotolo, Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes, *Chemistry and Biology* 16 (2009) 382–390.
- [15] A. Politis, A.Y. Park, S.J. Hyung, D. Barsky, B.T. Ruotolo, C.V. Robinson, Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes, *PLoS ONE* 5 (2010) e12080.
- [16] D.P. Smith, K. Giles, R.H. Bateman, S.E. Radford, A.E. Ashcroft, Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry, *Journal of the American Society for Mass Spectrometry* 18 (2007) 2180–2190.
- [17] D.P. Smith, S.E. Radford, A.E. Ashcroft, Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry, *Proceedings of the National Academy of Sciences of the United States of America* 107 (2010) 6794–6798.
- [18] C. Uetrecht, R.J. Rose, E. van Duijn, K. Lorenzen, A.J. Heck, Ion mobility mass spectrometry of proteins and protein assemblies, *Chemical Society Reviews* 39 (2010) 1633–1655.
- [19] D.P. Smith, L.A. Woods, S.E. Radford, A.E. Ashcroft, Structure and dynamics of oligomeric intermediates in  $\beta$ 2-microglobulin self-assembly, *Biophysical Journal* 101 (2011) 1238–1247.
- [20] C. Costa Mdo, H.L. Paulson, Toward understanding Machado-Joseph disease, *Progress in Neurobiology* 97 (2012) 239–257.
- [21] Y. Kawaguchi, T. Okamoto, M. Taniwaki, M. Aizawa, M. Inoue, S. Katayama, S. Kawakami, M. Nishimura, I. Akiyoshi, CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1, *Nature Genetics* 8 (1994) 221–228.
- [22] C.A. Matos, S. de Macedo-Ribeiro, A.L. Carvalho, Polyglutamine diseases: the special case of ataxin-3 and Machado-Joseph disease, *Progress in Neurobiology* 95 (2011) 26–48.
- [23] H.L. Paulson, Machado-Joseph disease/spinocerebellar ataxia type 3, *Handbook of Clinical Neurology* 103 (2012) 437–449.
- [24] T.S. Takahashi, S. Katada, O. Onodera, Polyglutamine diseases: where does toxicity come from? What is toxicity? Where are we going? *Journal of Molecular and Cellular Biology* 2 (2010) 180–191.
- [25] R. Wetzel, Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence, *Journal of Molecular Biology* 421 (2012) 466–490.
- [26] L. Masino, V. Musi, R.P. Menon, P. Fusi, G. Kelly, T.A. Frenkiel, Y. Trotter, A. Pastore, Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail, *FEBS Letters* 549 (2003) 21–25.
- [27] X.H. He, F. Lin, Z.H. Qin, Current understanding on the pathogenesis of polyglutamine diseases, *Neuroscience Bulletin* 26 (2010) 247–256.
- [28] A. Vitalis, X. Wang, R.V. Pappu, Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories, *Biophysical Journal* 93 (2007) 1923–1937.
- [29] G. Nicastro, R.P. Menon, L. Masino, P.P. Knowles, N.Q. McDonald, A. Pastore, The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition, *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005) 10493–10498.
- [30] C.J. Cummings, H.Y. Zoghbi, Fourteen and counting: unraveling trinucleotide repeat diseases, *Human Molecular Genetics* 9 (2000) 909–916.
- [31] G.M. Harris, K. Dodelzon, P. Gong, H.L. Gonzalez-Alegre, H.L. Paulson, Splice isoforms of the polyglutamine disease protein ataxin-3 exhibit similar enzymatic yet different aggregation properties, *PLoS ONE* 5 (2010) e13695.

- [32] L. Masino, G. Nicastro, R.P. Menon, F. Dal Piaz, L. Calder, A. Pastore, Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3, *Journal of Molecular Biology* 344 (2004) 1021–1035.
- [33] G. Nicastro, M. Habeck, L. Masino, D.I. Svergun, A. Pastore, Structure validation of the Josephin domain of ataxin-3: conclusive evidence for an open conformation, *Journal of Biomolecular NMR* 36 (2006) 267–277.
- [34] Y. Mao, F. Senic-Matuglia, P.P. Di Fiore, S. Polo, M.E. Hodson, P. De Camilli, Deubiquitinating function of ataxin-3: insights from the solution structure of the Josephin domain, *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005) 12700–12705.
- [35] C. Santambrogio, A.M. Frana, A. Natalello, E. Papaleo, M.E. Regonesi, S.M. Doglia, P. Tortora, G. Invernizzi, R. Grandori, The role of the central flexible region on the aggregation and conformational properties of human ataxin-3, *The FEBS Journal* 279 (2012) 451–463.
- [36] L. Gales, L. Cortes, C. Almeida, C.V. Melo, M.C. Costa, P. Maciel, D.T. Clarke, A.M. Damas, S. Macedo-Ribeiro, Towards a structural understanding of the fibrillization pathway in Machado-Joseph's disease: trapping early oligomers of non-expanded ataxin-3, *Journal of Molecular Biology* 353 (2005) 642–654.
- [37] P.G. Blommel, B.G. Fox, A combined approach to improving large-scale production of tobacco etch virus protease, *Protein Expression and Purification* 55 (2007) 53–68.
- [38] M.F. Mesleh, J.M. Hunter, A.A. Scvartsburg, G.C. Schatz, M.F. Jarrold, Structural information from ion mobility measurements: effects of the long-range potential, *Journal of Physical Chemistry* 100 (1996) 16082–16086.
- [39] C.T. Bleiholder, T. Wyttenback, M.T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method, *International Journal of Mass Spectrometry* 308 (2011) 1–10.
- [40] L. Testa, S. Brocca, R. Grandori, Charge-surface correlation in electrospray ionization of folded and unfolded proteins, *Analytical Chemistry* 83 (2011) 6459–6463.
- [41] A.K. Frimpong, R.R. Abzalimov, V.N. Uversky, I.A. Kaltashov, Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of  $\alpha$ -synuclein, *Proteins* 79 (2010) 714–722.
- [42] J. Pan, L. Konermann, Calcium-induced structural transitions of the calmodulin–melittin system studied by electrospray mass spectrometry: conformational subpopulations and metal-unsaturated intermediates, *Biochemistry* 49 (2010) 3477–3486.
- [43] C.S. Hoaglund-Hyzer, A.E. Counterman, D.E. Clemmer, Anhydrous protein ions, *Chemical Reviews* 99 (1999) 3037–3080.
- [44] J. Li, J.A. Taraszka, A.E. Counterman, D.E. Clemmer, Influence of solvent composition and capillary temperature on the conformations of electrosprayed ions: unfolding of compact ubiquitin conformers from pseudonative and denatured solutions, *International Journal of Mass Spectrometry* 185–187 (1999) 37–47.
- [45] S.J. Valentine, J.G. Anderson, A.D. Ellington, D.E. Clemmer, Disulfide-intact and -reduced lysozyme in the gas phase: conformations and pathways of folding and unfolding, *Journal of Physical Chemistry B* 101 (1997) 3891–3900.
- [46] A.H. Mao, S.L. Crick, A. Vitalis, C. Chicoine, R.V. Pappu, Net charge per residue modulates conformational ensembles of intrinsically disordered proteins, *Proceedings of the National Academy of Sciences of the United States of America* 107 (2010) 8183–8188.
- [47] L. Masino, G. Nicastro, A. De Simone, L. Calder, J. Molloy, A. Pastore, The Josephin domain determines the morphological and mechanical properties of ataxin-3 fibrils, *Biophysical Journal* 100 (2011) 2033–2042.
- [48] L. Masino, G. Nicastro, L. Calder, M. Vendruscolo, A. Pastore, Functional interactions as a survival strategy against abnormal aggregation, *Journal of the Federation of the American Societies for Experimental Biology* 25 (2011) 45–54.